State of Delaware
Final Report: Ozone Observations and Forecasts in 2014

A Report Prepared for the Delaware Department of Natural Resources and Environment

Dr. Amy K. Huff
William F. Ryan
Department of Meteorology
The Pennsylvania State University

February 27, 2015

Table of Contents

List of Figures

Executive Summary

- The O_3 season of 2014 was the second consecutive historically low O_3 season for the State of Delaware. Only 3 days in 2014 reached the Code Orange threshold, and there were no Code Red or Code Purple days.
- Two seasons in a row are insufficient to determine if the historically low O_3 conditions in 2013-2014 are a trend or an anomaly. The clean conditions during the summer of 2014 are attributed to continued reductions in O_3 precursor emissions, slightly cooler weather than average, and an atypical weather pattern that limited the development of heat waves and the frequency of westerly transport, which are historically associated with poor air quality days in the Mid-Atlantic region.
- Overall forecast skill (all days) was similar to recent years, with a median absolute error of 6.0 ppbv (compared to 6.5 ppbv in 2013). As in 2013, absolute errors were slightly higher than in 2011-2012 (5.0 ppbv) due to the challenge of forecasting during the historically clean O_3 season.
- As a consequence of the historically clean O_3 season, forecast skill for the Code Orange O_3 cases was poor. The false alarm rate in 2014 (0.75) was substantially higher than 2012 (0.32) and 2011 (0.43), but less than 2013 (1.00). The total number of false alarms (3) in 2014 continued to drop, however, down from 4 in 2013, 6 in 2012 and 10 in 2011. The forecast hit rate (0.33) was also low, but better than 2013 (0.00).
- Historically, hot weather (maximum temperature $\geq 90°$ F) has been strongly associated with and high O_3. Over the past 5 years, this relationship has substantially weakened, and this trend continued in 2014. Temperature alone is no longer a necessary *and* sufficient predictor of Code Orange or higher O_3.
- Persistence and regional O_3 concentrations were not consistently reliable forecast factors in 2014, increasing the difficulty of accurately forecasting Code Orange cases.
- As in 2013, the numerical air quality models that provide forecast guidance had similar poor skill for the Code Orange O_3 cases and a propensity to over-predict peak O_3 in all cases. The NOAA-EPA model had the most skillful performance overall of the three forecast models used in the model "ensemble."
- An updated statistical model using 2008-2013 observations has been created and will be tested as operational guidance in Delaware during the 2015 O_3 season.

Ozone Observations in 2014

The ozone (O_3) season of 2014 was the second historically low O_3 season in a row for Delaware. As shown in Figure 1, the 2014 seasonal (May 1-September 30) mean and median peak 8-hour observed O_3 concentrations were the second-lowest since 1990; only the 2013 values were lower. The daily time series of peak 8-hour observed O_3 in Delaware during the summer of 2014 compared to the 2003-2013 average (Figure 2) demonstrates that daily peak 8-hour observed O_3 concentrations in 2014 fell below the 2003-2013 average on most days. This difference is further illustrated in Figure 3, the daily time series of the difference between peak 8-hour observed O_3 concentrations in 2014 compared to the 2003-2013 average.

Similar to 2013, the 2014 O_3 season also had a remarkably low occurrence of high O_3 days. In 2014, there were only 3 days with peak 8-hour observed O_3 in excess of the Code Orange threshold of 76 ppbv, no days in exceedance of the Code Red threshold of 96 ppbv, and no days in exceedance of the Code Purple threshold of 116 ppbv (Figure 4). Thus, 2014 ties 2009 for the second-lowest occurrence of high O_3 days since 1990, just slightly above the total of 2 Code Orange O_3 days in 2013. Code Orange O_3 days in 2014 were well below the average frequency of 16 days per year for the 2003-2013 period. In addition, there were no truly extreme O_3 days (8-hour $O_3 \geq 105$ ppbv) in 2014. The last year that 8-hour $O_3 \geq 105$ ppbv was observed was 2012, and the last year that 8-hour $O_3 \geq 115$ ppbv was observed was 2007.

The recent downward trends in seasonal mean and median peak 8-hour observed O_3 concentrations (Figure 1) and incidents of high O_3 days (8-hour $O_3 \geq 76$ ppbv; Figure 4) are partly attributable to continued regional reductions in O_3 precursor emissions. These regional reductions began with the "NO_x SIP Call" that went into effect circa 2002. Thus, the years since 2002 have been marked by reductions in O_3 concentrations across the Mid-Atlantic region. In Delaware, observed Code Orange O_3 days have decreased by 60%, observed Code Red O_3 days have decreased by 88%, and observed Code Purple O_3 days have decreased by 90% for the period 2003-2014 compared to 1990-2002. Pie charts showing changes in the frequency of all observed AQI color codes since 1990 further illustrate this trend (Figures 5-7). The very high percentage of Code Green days observed in 2014 (82%) compared to 2003-2013 (60%) and 1990-2002 (46%) underlines the very clean O_3 conditions that occurred in 2014.

The Impact of Meteorology in 2014

The meteorology of summer 2014 also played a contributing role in the historically low observed O_3 concentrations. Temperatures in Delaware during June-August 2014 were slightly below (0.05 °F) the long-term average (Figure 8). Because O_3 has a photochemical source, hot weather (maximum temperature ≥ 90° F) and high O_3 (Code Orange or higher) are correlated, although this relationship has been weakening in recent years in the Mid-Atlantic region. Thus, slightly below average temperatures in the peak of summer 2014 likely helped to suppress O_3 formation. Figure 9 shows that precipitation in Delaware during June-August 2014 was approximately the same as the long-term average. Cloud cover associated with precipitation blocks sunlight and inhibits O_3 formation. Since 2014 was not an unusually wet summer, however, precipitation was likely not a contributing factor to the historically low observed O_3 concentrations.

The weather pattern during summer 2014 was very progressive, meaning that storm systems and cold fronts moved through the Mid-Atlantic on a regular basis. Consequently, there were no extended periods of hot weather or heat waves (3 or more days in a row with maximum temperatures ≥ 90° F). During summers with average or above average O_3 concentrations, the semi-permanent Bermuda High extends westward over the Mid-Atlantic region, allowing extended periods of hot, dry, and stagnant weather to develop, which are conducive for O_3 formation. During summer 2014, the Bermuda High stayed well to the east (Figure 10) and was suppressed southward, which helped to prevent the development of heat waves across the Mid-Atlantic. With the Bermuda High circulation suppressed southward, more frequent pulses of cool air reached the Mid-Atlantic. Figure 11 shows the below average 1000-500 mb thicknesses during July 2014, which is indicative of the cooler than average air over the eastern U.S. This weather pattern was not favorable for O_3 formation, and it was likely a contributing factor to the historically low observed O_3 concentrations in 2014.

As in 2013, another key factor in limiting peak O_3 concentrations during summer 2014 appeared to be the predominant transport pattern aloft. During summer 2014, the transport pattern differed from the classic westerly transport pattern, in which air aloft at approximately 500-1500 m above ground level (AGL) flows from the Ohio River Valley, a source region for O_3 and O_3 precursors, such as NO_x. The summer 2014 transport pattern featured stronger than usual southerly flow along and east of I-95, which brought clean maritime air into Delaware and the Mid-Atlantic region, as shown in Figure 12. Winds in the transport layer aloft were suppressed and shifted northerly compared to normal, with the result that large scale westerly transport from the Ohio River Valley and the Midwest was limited.

Code Orange O_3 Concentrations in 2014

The three observed Code Orange O_3 days in 2014 were single-day events that occurred in early July (July 8 and 11) and late August (August 27). Temperature and O_3 are strongly correlated overall, but the association of hot weather with Code Orange or higher O_3 has been steadily weakening in Delaware since 2002, and this trend continued in 2014. Only one of the Code Orange days (July 8) coincided with hot weather, when maximum temperature was 92 °F at Dover, Delaware. The other two days had maximum temperatures of 85 °F (July 11) and 89 °F (August 27). The percent of days exceeding the Code Orange threshold for a given range of maximum surface temperature in Delaware from 1990-2014 is presented in Figure 13. While days with maximum surface temperature ≥ 95° F were historically very likely to be poor air quality days, the frequency of Code Orange or higher cases on slightly less hot days (90-94 °F) has dropped from 70% prior to 2003 to 29% in 2003-2013 to only 6% (1 case) in 2014.

In addition to the weakening of the relationship between hot weather and high O_3, Code Orange or higher days are becoming less likely to occur in multi-day events. Since 2003, almost 50% or more of the observed Code Orange or higher O_3 cases occurred in multi-day events (2 or more days in a row), as shown in Figure 14. The exceptions were the historically low O_3 years of 2009, 2013, and 2014. It seems that in years when O_3 concentrations are low overall, high O_3 cases occur in single day "spikes" and not in multi-day events, as is more typical in years when observed O_3 concentrations are average or above average. This shift away from multi-day events for 2013-2014 may be an anomaly related to atypical weather patterns or it may be the beginning of a longer trend due to continuing reductions in O_3 precursor emissions.

Skill of Ozone Forecasts in 2014

The skill of all O_3 forecasts in 2014 was comparable to recent years, although absolute errors were slightly higher than in 2011 and 2012 due to the difficulty of forecasting for the unusually clean O_3 season. A time series of forecasts and observations for 2014 is shown in Figure 15 and error statistics for forecasts during recent years (2011-2014) are given in Figure 16. Median absolute forecast error during the 2014 season was 6.0 ppbv, which is slightly higher than 2011 and 2012 (5.0 ppbv) but lower than 2013 (6.5 ppbv). Figure 16 also shows that, similar to recent years, there was a bias toward over-prediction of O_3 in 2014 (2.6 ppbv), which is analogous to the over-prediction bias that occurred in 2011 (2.6 ppbv) and 2013 (2.5 ppbv).

As was the case in 2013, the 2014 forecasts were below average in skill for the Code Orange cases (Figure 17), although an improvement on 2013 skill. Details on the calculation of skill scores are given in Appendix A. The false alarm rate in 2014 was 0.75, indicating that for 3 of the 4 forecasted Code Orange days, Code Orange O_3 was not observed. One of these days (June 17) was a close miss, with observed O_3 of 73 ppbv. The other two days were substantial misses, however, with observed O_3 of only 62 ppbv on July 12 and 64 ppbv on August 5. Single day "spikes" in O_3 are difficult to forecast in part because persistence, or previous day peak 8-hour O_3, is an important predictor. Single day Code Orange "spikes" have been even more difficult to forecast in recent years as the relationship between hot weather and O_3 has become more tenuous, as noted above. The recent downward trend in the total number of false alarms continued in 2014, with a total of 3, compared to 4 in 2013, 6 in 2012, and 10 in 20111. The hit rate, the ratio of Code Orange forecasts to observed Code Orange, was 0.33, meaning that only 1 of 3 observed Code Orange days (August 27) was correctly forecasted with health alerts issued to the public. This is an improvement on the hit rate of 0.00 in 2013. The very low occurrence of observed Code Orange O_3 days in 2014 is primarily responsible for the low skill compared to 2011 and 2012; there were only 3 observed Code Orange days in 2014 compared to 14 in 2011 and 20 in 2012.

In 2014, the majority (2) of Code Orange cases in occurred in sub-90 ° F conditions, making it substantially more difficult to distinguish accurately conditions conducive to Code Orange O_3. In the past, forecasters could rely on the strong relationship between hot weather and high O_3 to help predict Code Orange or higher conditions. However, temperature alone is no longer a necessary *and* sufficient predictor of Code Orange O_3. The "decoupling" of O_3 and temperature means that other conditions, such as local scale wind circulation, convection, cloud cover, and local emissions, are becoming greater contributing factors for observed O_3 concentrations. These phenomena, which occur on scales ("mesoscale") that are too small for current numerical air quality forecast models to resolve, are more difficult to forecast, which in turn increases the difficulty of forecasting Code Orange O_3 days. In addition, persistence and regional build-up of high O_3 have historically been very reliable factors for predicting Code Orange or higher O_3. But in the absence of multi-day events in 2014, persistence and regional O_3 concentrations were not predictive of the observed Code Orange cases. In fact, for the July 11 Code Orange case, the previous day's peak 8-hour observed O_3 was 57 ppbv, which is in the Code Green range. Thus, based on persistence, observed O_3 on July 11 should have been in the Code Green or at most the Code Yellow range; Code Orange was completely unexpected. The numerical ozone forecast models are not accurately predicting Code Orange or higher O_3 under these conditions either, as discussed in the next section. This leaves the forecaster with few tools on which to rely. To fill this gap, a statistical O_3 model for Delaware has been updated based on observations from 2008-2013, as discussed in the "Outlook for 2015" section.

Performance of Ozone Numerical Forecast Models in 2014

Numerical air quality forecast models are gradually becoming able to provide reliable forecast guidance for operational O_3 forecasts. These models become operational in the mid-2000s and have steadily improved in skill since that time. For the 2011-2012 O_3 seasons, a number of numerical air quality forecast models were routinely available to forecasters. The most skillful models used by operational forecasters in previous years include the NOAA-EPA air quality model (NOAA), the SUNY-Albany model (SUNY) and two versions of the Baron Advanced Meteorological Services model (BAMS), the CMAQ and RT. In 2013, the SUNY model was not operational and the NOAA model came close to cancellation due to federal budget reductions. As a result, the annual updates to the NOAA model for emissions and other components were not made in 2013. Because the SUNY model, which was relatively accurate in 2011-2012, was not available in 2013, the North Carolina Department of Environment and Natural Resources (NCDENR) O_3 model was utilized instead. In 2014, the SUNY model was still unavailable. NOAA resumed support for the NOAA-EPA model in 2014 and made updates to the emissions and chemistry. Thus, the model "ensemble" for 2014 included the NOAA-EPA model, the BAMS CMAQ and RT, and the NCDENR model.

Error statistics for the O_3 forecast models used as guidance in 2014 are shown in Figure 18, and skill scores for the Code Orange O_3 predictions are given in Figure 19. All of the forecast models suffered from an over-prediction bias, ranging from 1.4 ppbv (NOAA-EPA) to 4.6 ppbv (NCDENR). The median absolute error of the models varied from 5.6 ppbv (BAMS-RT) to 7.0 ppbv (NCDENR). All of the models had the same hit rate of 0.33, indicating that they accurately predicted 1 of the 3 observed Code Orange days. For all of the models, this day was August 27, which is the same day that the expert forecaster correctly forecasted Code Orange. This underscores the potential usefulness of numerical forecast model guidance for operational forecasts.

For the second year in a row, the most skillful model overall was the NOAA-EPA model. It had the lowest bias of all the models by 2.0 ppbv, and the second lowest median absolute error, just behind the BAMS-RT. The NOAA-EPA model was the only model in the ensemble that did not issue a false alarm for Code Orange O_3, compared to 2 false alarms for all of the other models. Somewhat unexpectedly, the ensemble average forecasts were not more skillful than the NOAA-EPA model, or any of the other models, as is typically the case. The relatively higher median absolute error contribution from the NCDENR model appears to have skewed the accuracy of the ensemble forecast for 2014. This underscores the need for continuous verification of the model guidance during the course of the summer season, so variations in model performance can be determined and the most skillful model can be weighted more heavily as operational guidance. This approach will continue for the 2015 summer season.

Outlook for 2015

A new statistical model, trained on the period from 2008-2013, has been developed for Delaware. Statistical guidance had been discontinued in Delaware because changes in emissions of O_3 precursors associated with the NO_x SIP Rule, occurring in the period 2002-2007, significantly reduced O_3 across the Mid-Atlantic region. As a result, the assumption of "stationarity" for the model, the stability in the relationships between predictors and predictands, was violated and the model was no longer useful. Statistical tests indicate that the period 2008-2013 provides a more stable training set, so the expectation is that a model trained on these data will prove skillful. The updated statistical model will be tested in 2015, with the hope that it will help make Code Orange and higher O_3 days easier to identify. The model relies heavily on temperature and persistence, however, which may limit its usefulness if 2015 is similar to 2013-2014, when temperature and persistence were unreliable forecast factors.

The U.S. Environmental Protection Agency (EPA) has proposed lowering the 8-hour O_3 National Ambient Air Quality Standard (NAAQS) from 75 to 70 ppbv. If this new lower standard comes into effect for summer 2015, the number of Code Orange days will increase, and it will likely be somewhat easier to identify Code Orange cases.

Summary

The 2014 O_3 season was the second consecutive historically low O_3 season for the State of Delaware. Only 3 days in 2014 reached the Code Orange threshold (compared to 2 in 2013), and there were no Code Red or Code Purple days. Two seasons in a row are not sufficient to determine if the historically low O_3 conditions in 2013-2014 are an anomaly or the beginning of a trend. The clean conditions during the summer of 2014 are attributed to continued reductions in O_3 precursor emissions, slightly cooler weather than average, and an atypical weather pattern that limited the development of heat waves and the frequency of westerly transport, which are historically associated with poor air quality days in the Mid-Atlantic region.

Overall forecast skill (all days) was similar to recent years. Forecast skill for the Code Orange O_3 cases was poor, however, as a consequence of the historically low O_3 levels. The false alarm rate for Code Orange forecasts in 2014 (0.75) was substantially higher than 2012 and 2011, but an improvement on 2013 (1.00). The recent downward trend in the total number of false alarms (3) in 2014 continued, down from a high of 10 in 2011. The forecast hit rate (0.33) for Code Orange cases was also low in 2014, but better than 2013 (0.00). As in 2013, the numerical air quality models that provide forecast guidance had similar poor skill for the Code Orange O_3 cases and a propensity to over-predict peak O_3 in all cases. The NOAA-EPA model had the most skillful performance overall of the three forecast models used in the model “ensemble.”

Historically, hot weather and high O_3 have been strongly associated. Over the past 5 years, this relationship has substantially weakened, and this trend continued in 2014. Temperature alone is no longer a necessary *and* sufficient predictor of Code Orange or higher O_3. In addition, persistence and regional O_3 concentrations were not consistently reliable forecast factors in 2014, increasing the difficulty of accurately forecasting Code Orange cases. In part to fill the made by the loss of hot weather, persistence, and regional O_3 as reliable forecasting tools, an updated statistical model using 2008-2013 observations has been created and will be tested as operational guidance in Delaware during the 2015 O_3 season.

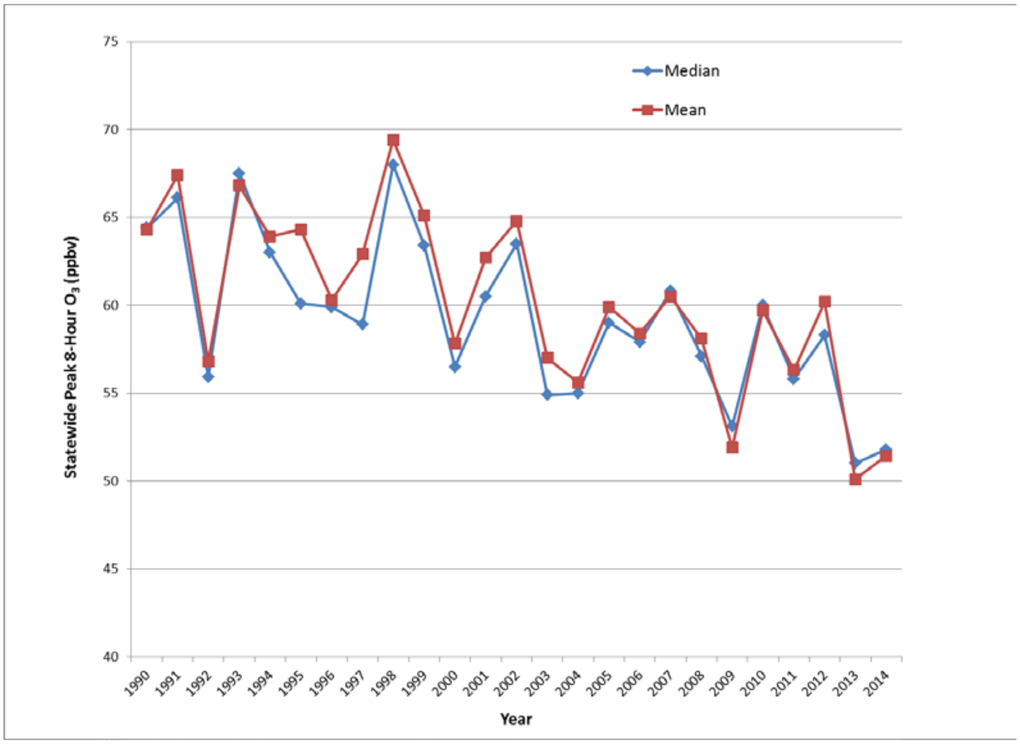

Figure 1. Seasonal (May-September) mean and median peak 8-hour observed O_3 in Delaware for 1990-2014.

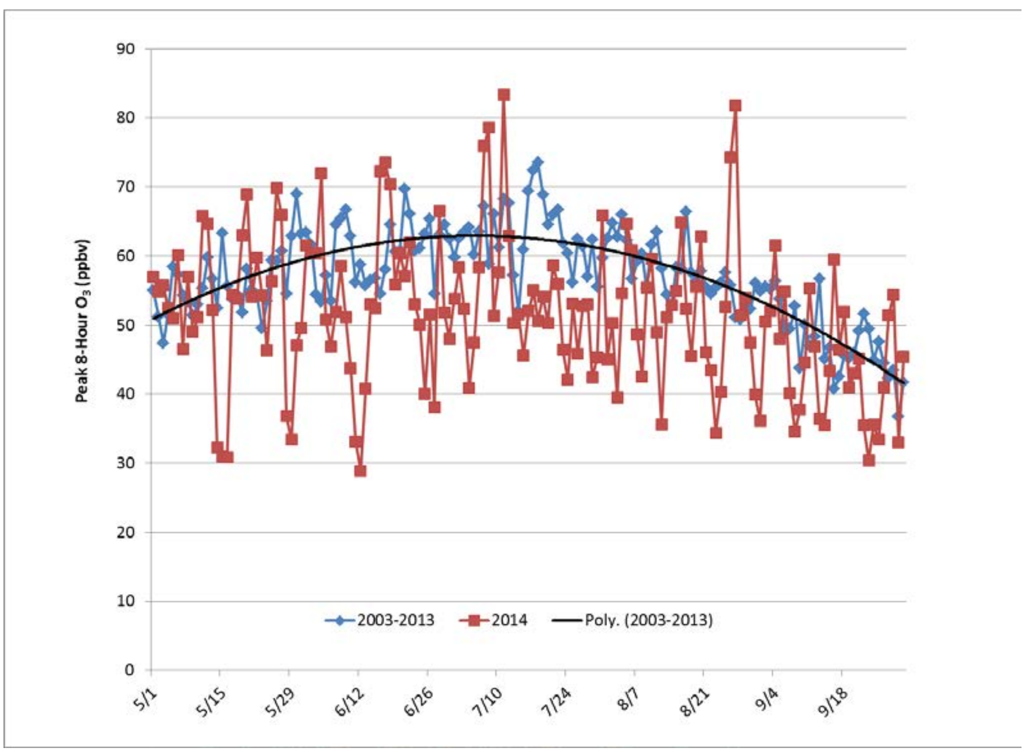

Figure 2. Daily time series of peak 8-hour observed O_3 in Delaware for 2014 (red line) compared to the 2003-2013 average (blue line). The black line is the best polynomial fit to the 2003-2013 average.

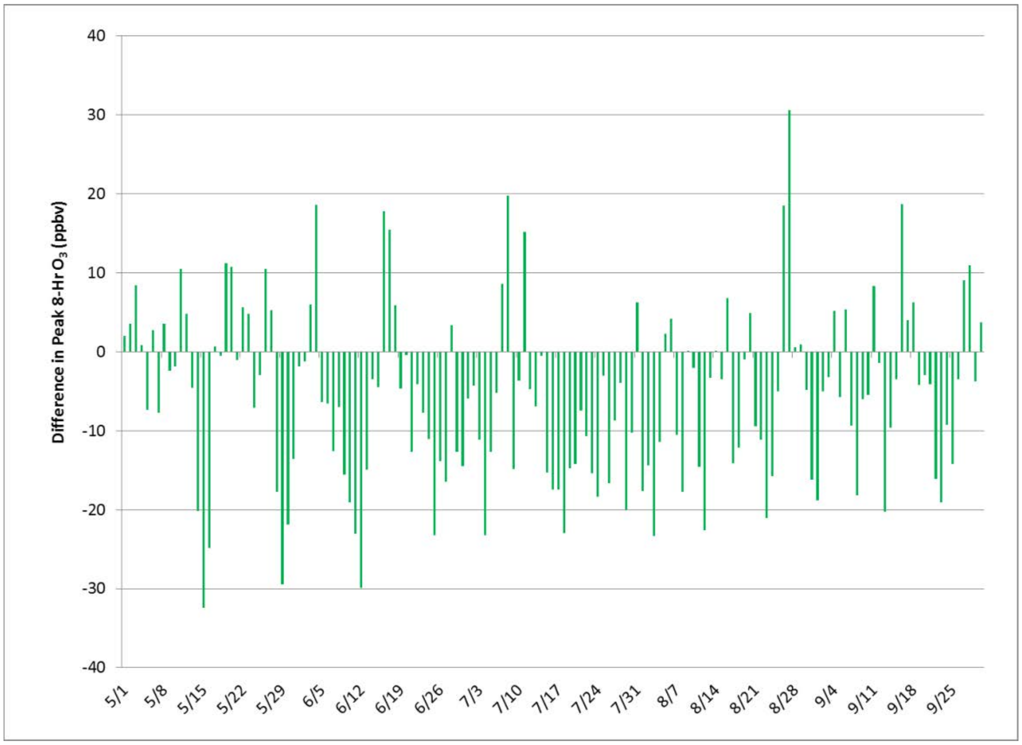

Figure 3. Daily time series of the difference between peak 8-hour observed O_3 for May 1 to September 30, 2014 compared to the 2003-2013 average.

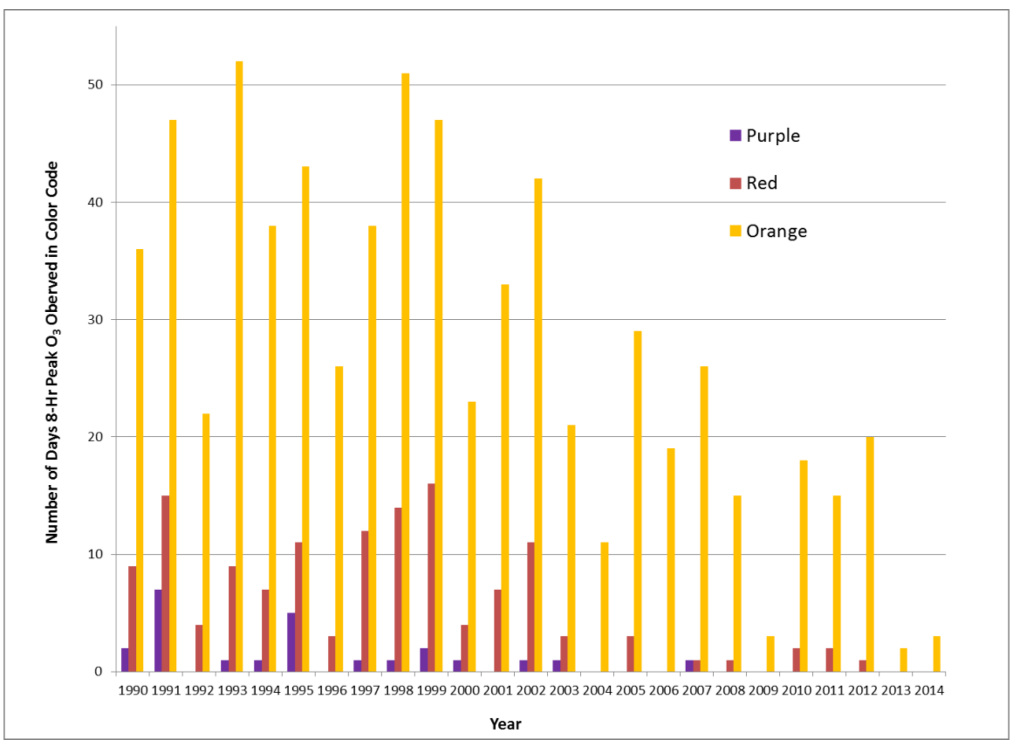

Figure 4. Frequency of Code Purple, Code Red, and Code Orange observed O_3 days in Delaware for 1990-2014.

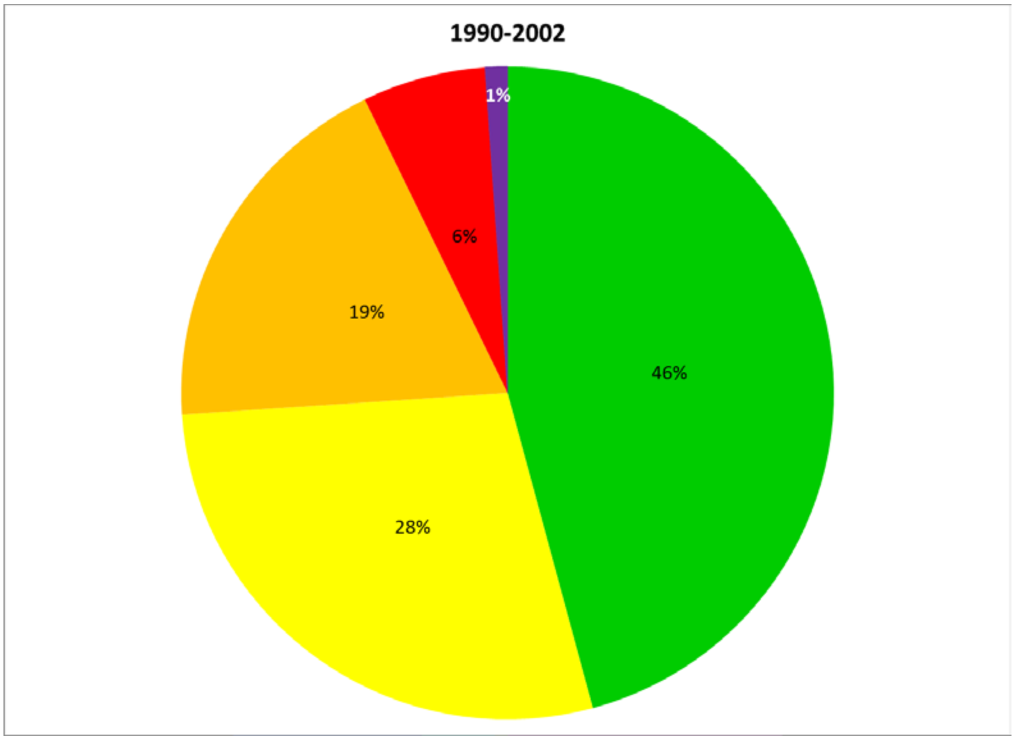

Figure 5. Frequency of AQI color codes for observed O_3 in Delaware for 1990-2002.

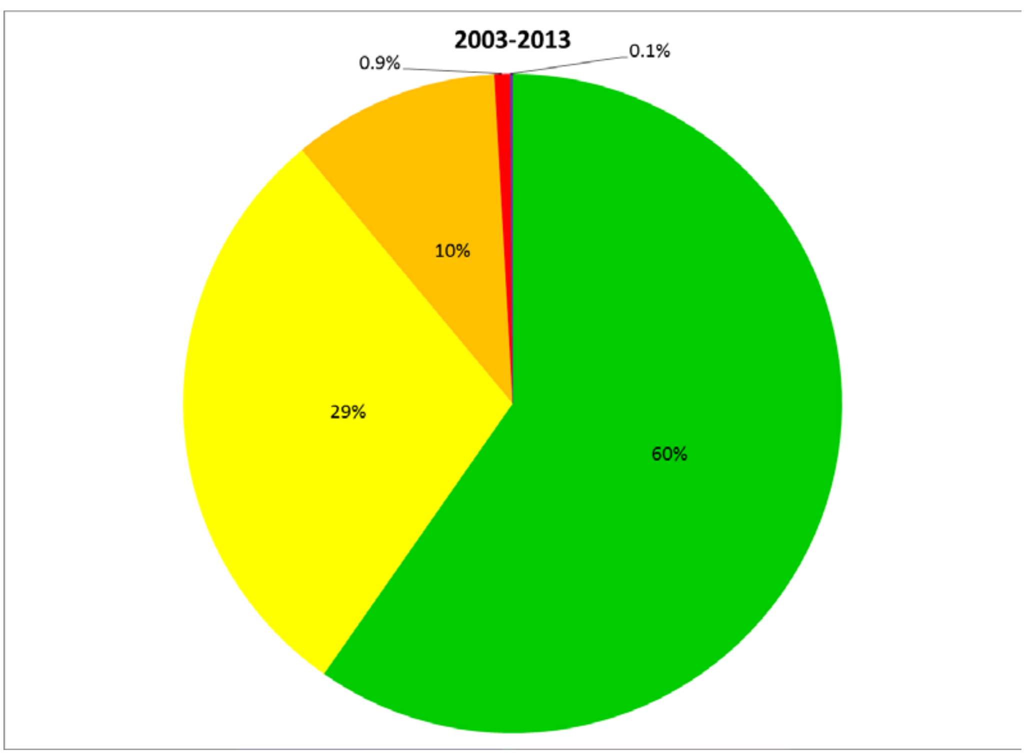

Figure 6. Frequency of AQI color codes for observed O_3 in Delaware for 2003-2013.

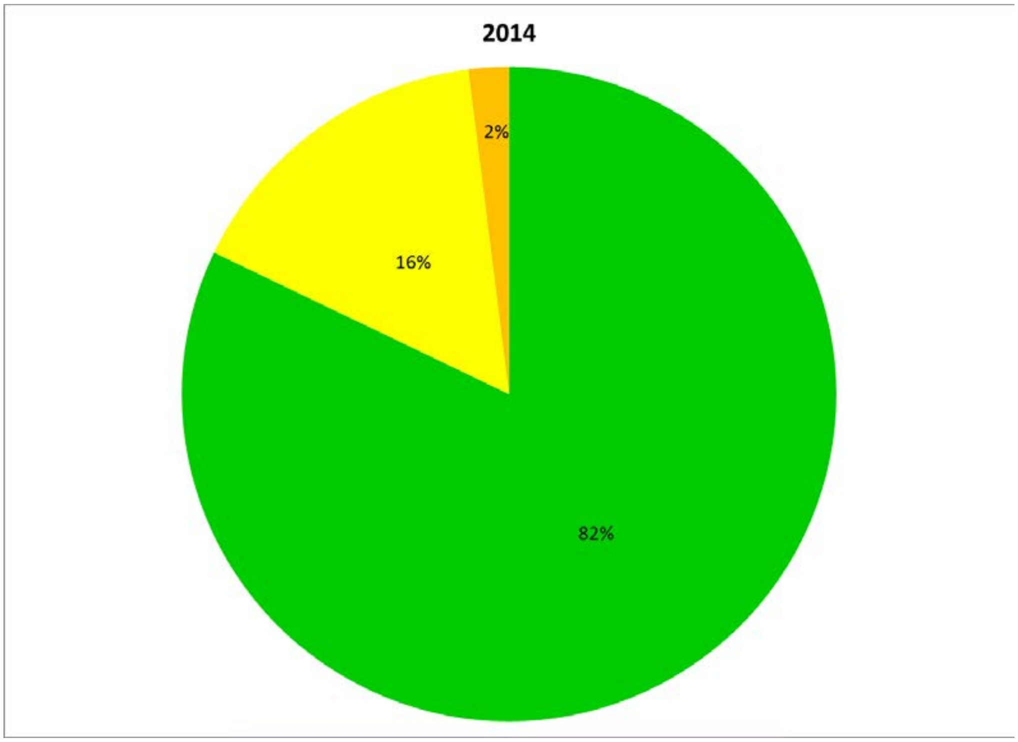

Figure 7. Frequency of AQI color codes for observed O_3 in Delaware for 2014.

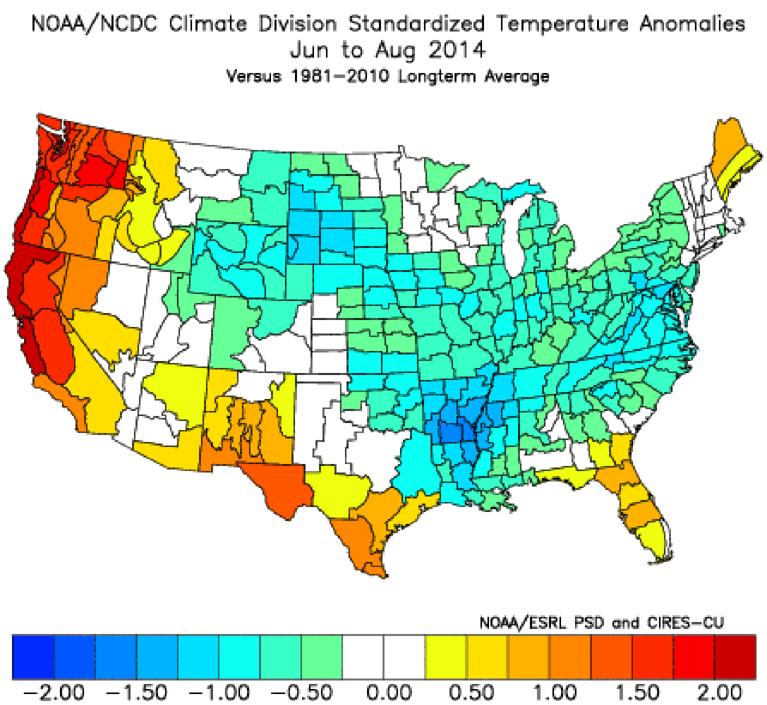

Figure 8. Temperature anomalies (in °F) in the U.S. for June-August 2014 compared to the 1981-2010 average (courtesy of NOAA/ESRL).

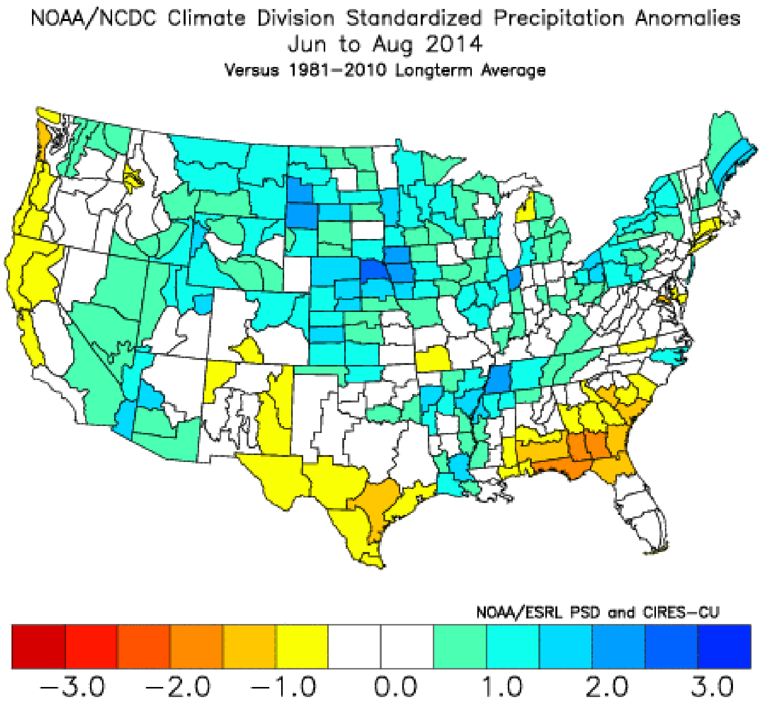

Figure 9. Precipitation anomalies (in inches) in the U.S. for June-August 2014 compared to the 1981-2010 average (courtesy of NOAA/ESRL).

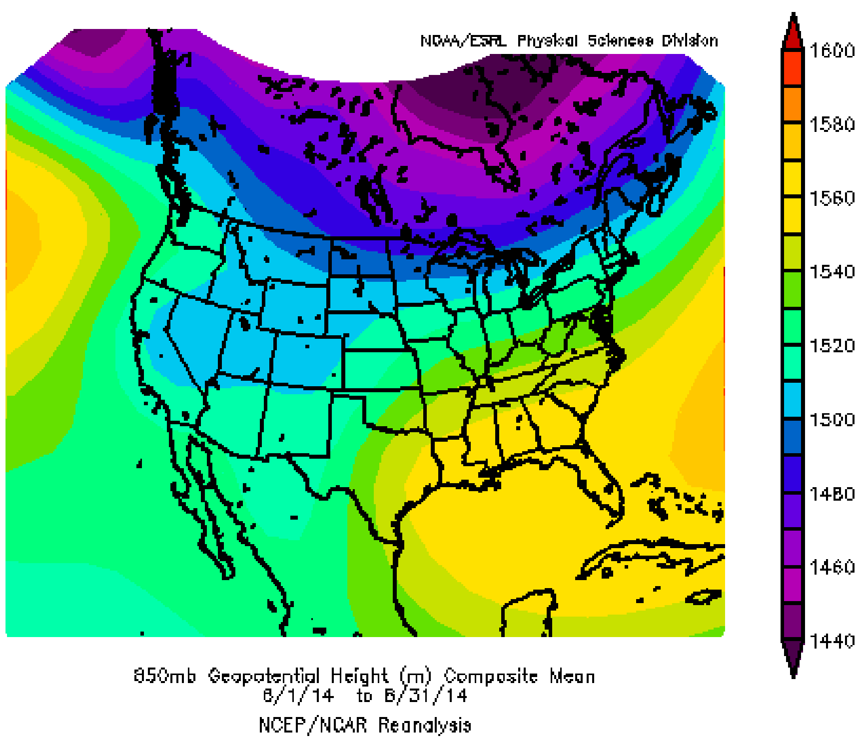

Figure 10. Geopotential height composite mean at 850 mb (~1500 m AGL) in the U.S. for June-August 2014 (courtesy NOAA/ESRL) showing the semi-permanent Bermuda High (yellow colors in Atlantic Ocean) suppressed southward and eastward compared to a summer with high O_3 observations.

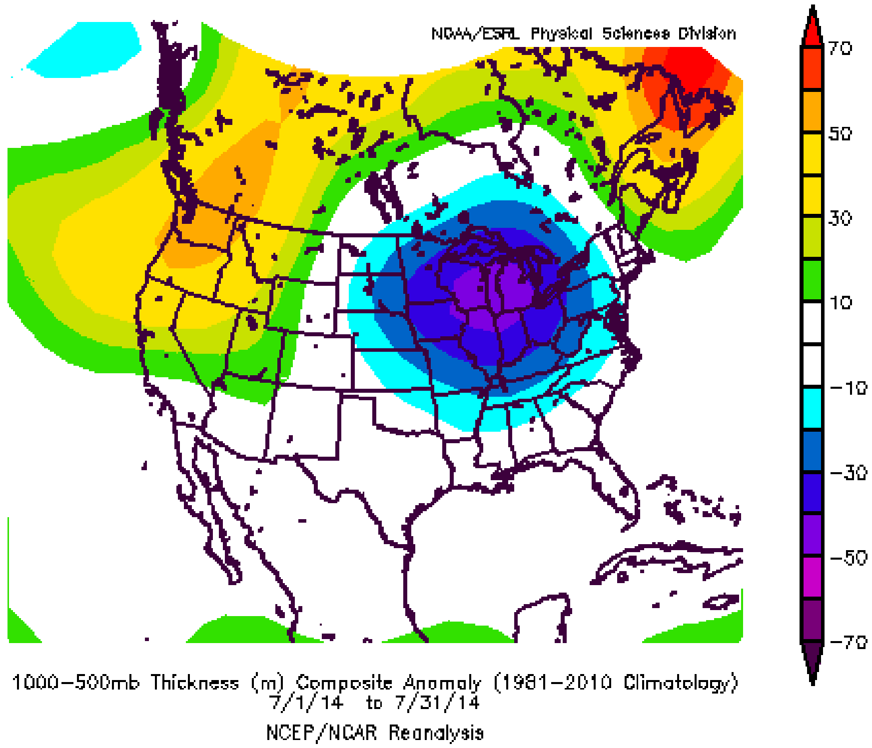

Figure 11. 1000-500 mb thickness composite anomaly in the U.S. for July 2014 (courtesy NOAA/ESRL).

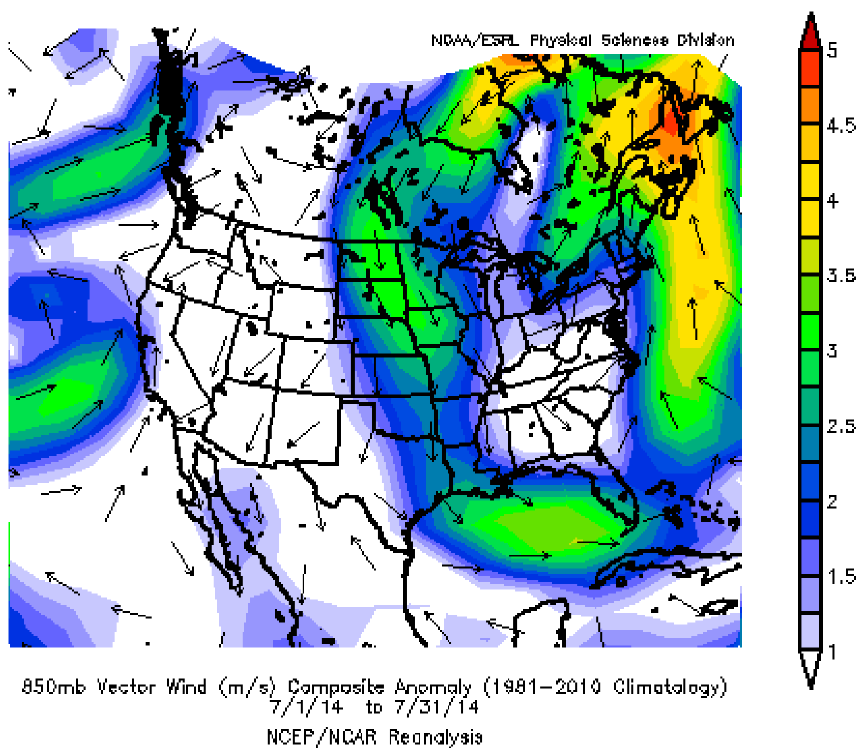

Figure 12. Vector wind composite anomaly at 850 mb (~1500 m AGL) in the U.S. for July 2014 (courtesy NOAA/ESRL).

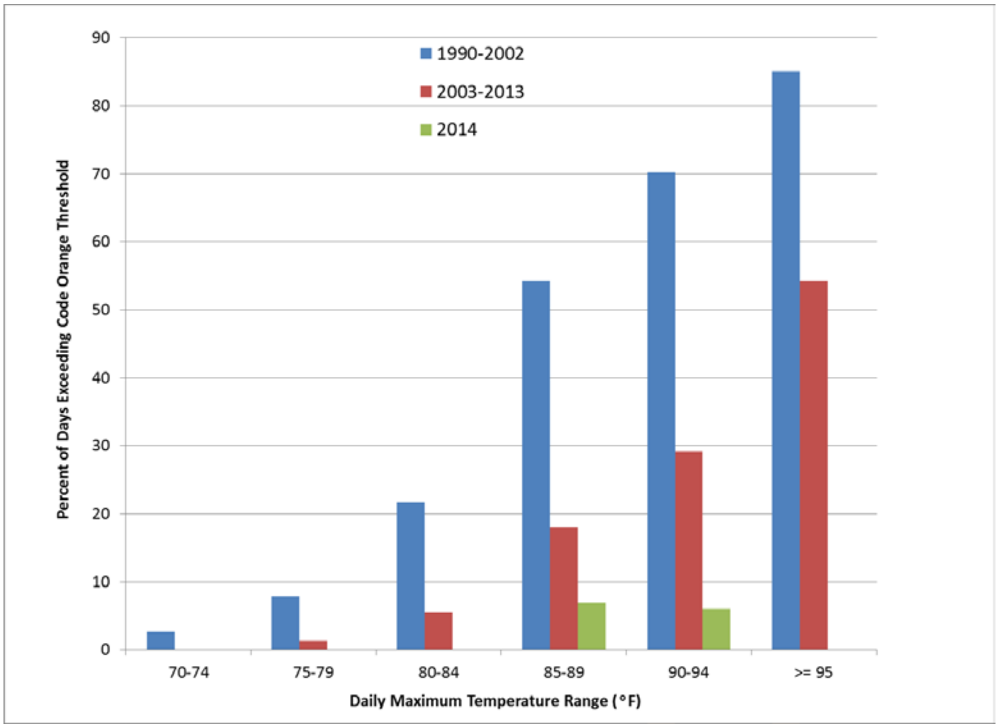

Figure 13. Percent of days exceeding the observed O_3 Code Orange threshold in Delaware for bins of maximum air temperature measured at Dover, Delaware (DOV).

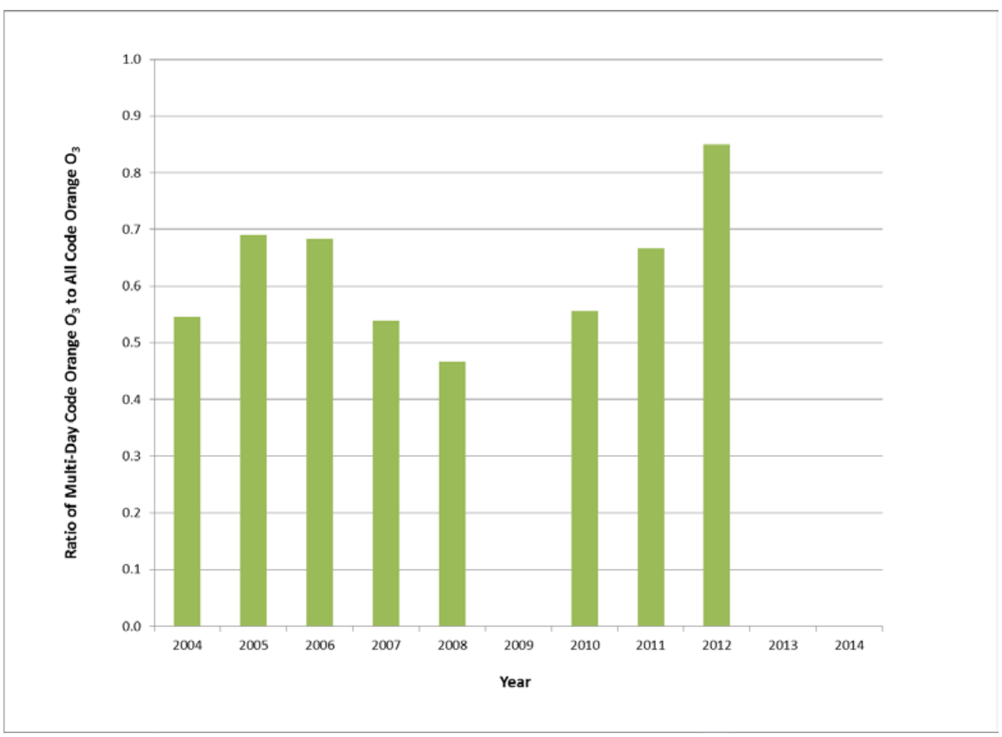

Figure 14. Ratio of multi-day (2 or more in a row) observed Code Orange or higher days to all Code Orange or higher days in Delaware for 2004-2014.

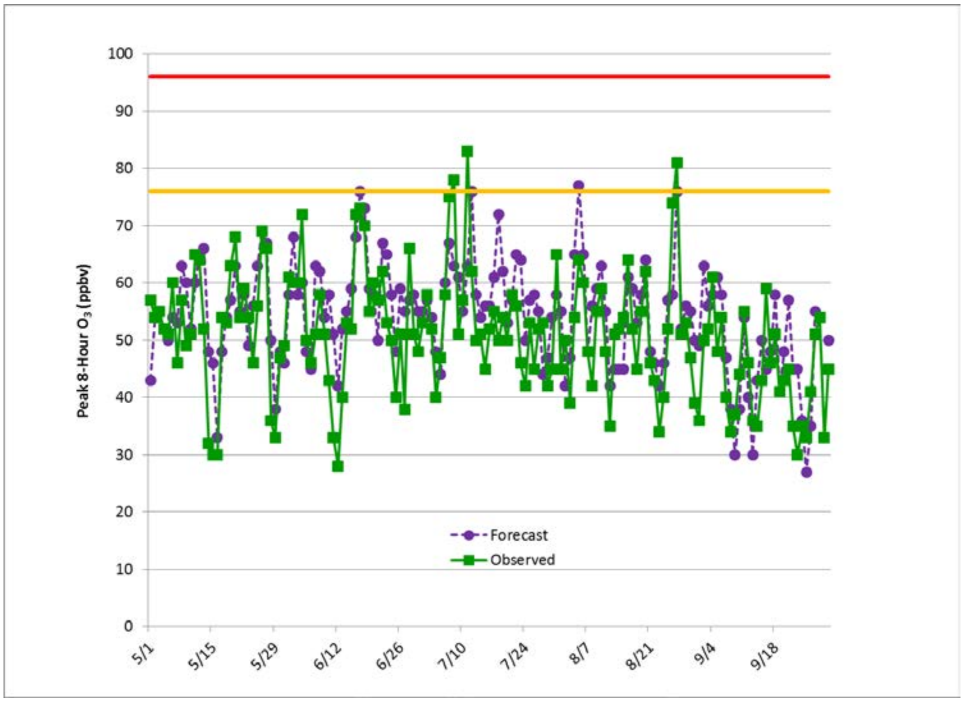

Figure 15. Peak 8-hour O_3 forecasts and observations for Delaware during May 1 to September 30, 2014. The orange and red lines indicate the Code Orange and Code Red thresholds, respectively.

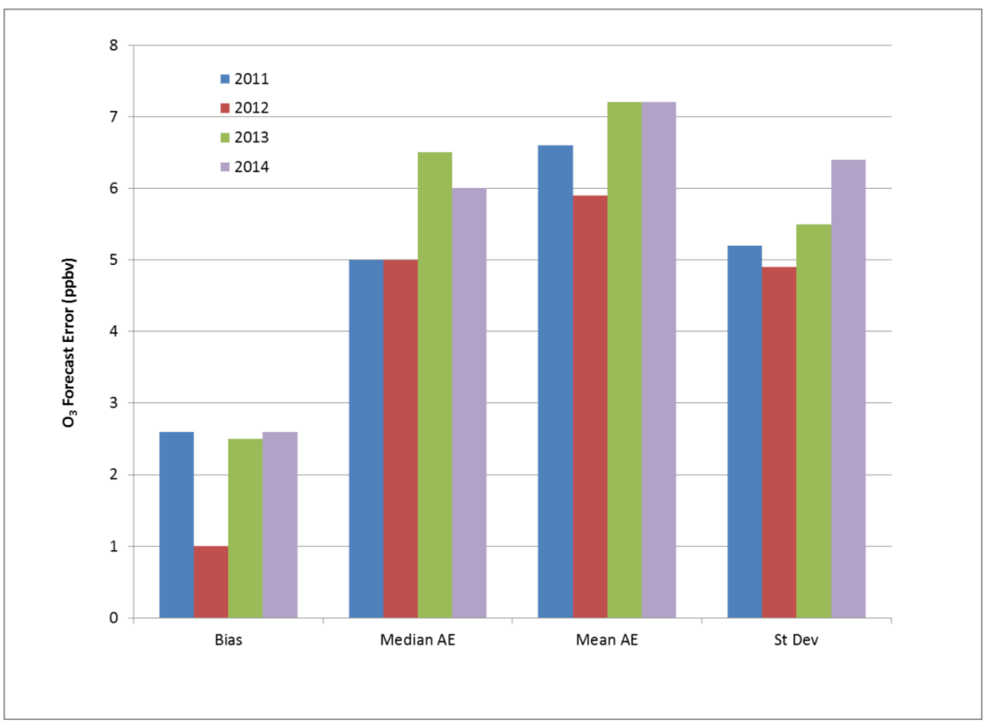

Figure 16. Error statistics for all peak 8-hour O_3 forecasts in Delaware for 2011-2014. "Median AE" refers to median absolute forecast error, "Mean AE" refers to mean absolute error, and "StDev" refers to the standard deviation of the mean absolute error.

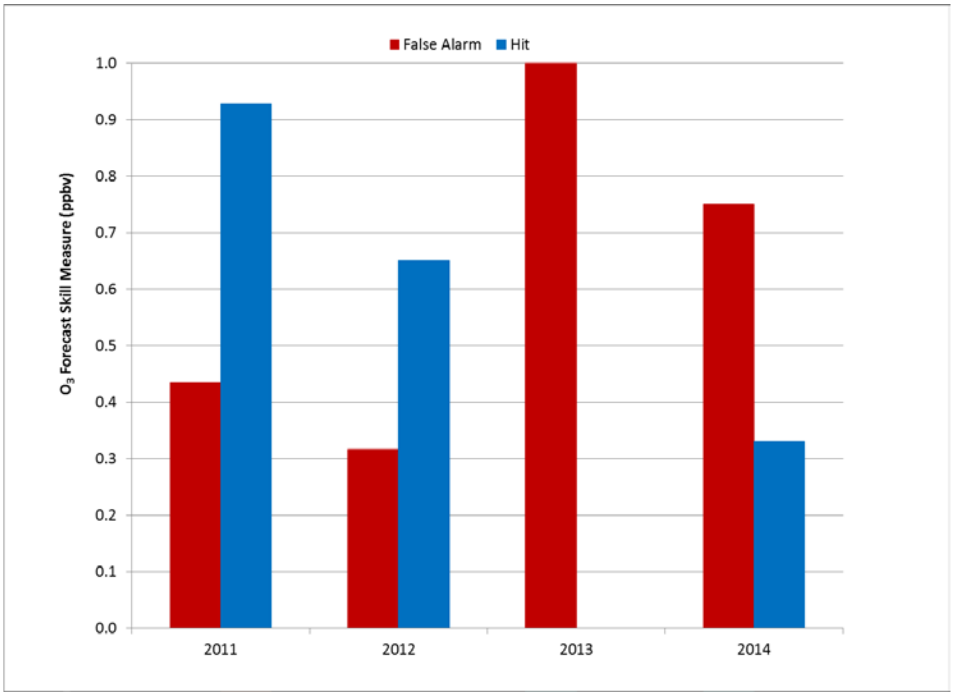

Figure 17. False alarm rate and hit rate for Code Orange O_3 forecasts in Delaware for 2011-2014.

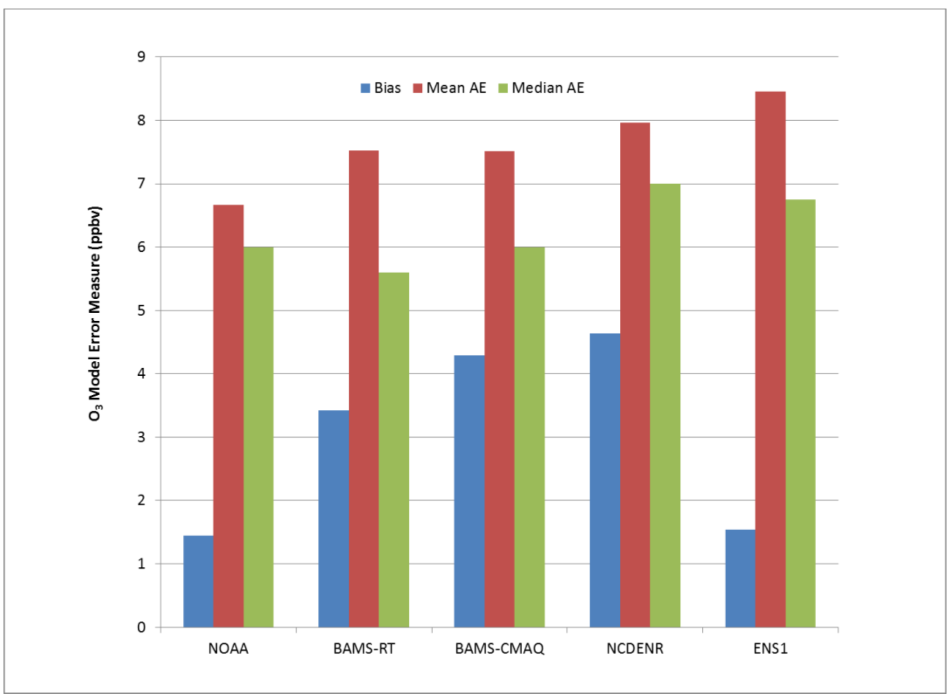

Figure 18. Error statistics (as in Figure 16) for air quality numerical forecast model guidance for all peak 8-hour O_3 predictions in Delaware in 2013. Two variations of the Baron Meteorological Services (BAMS) models are shown (CMAQ and RT). The ENS1 is a mean value from the NOAA, BAMS-RT, BAMS-CMAQ, and NCDENR model forecasts.

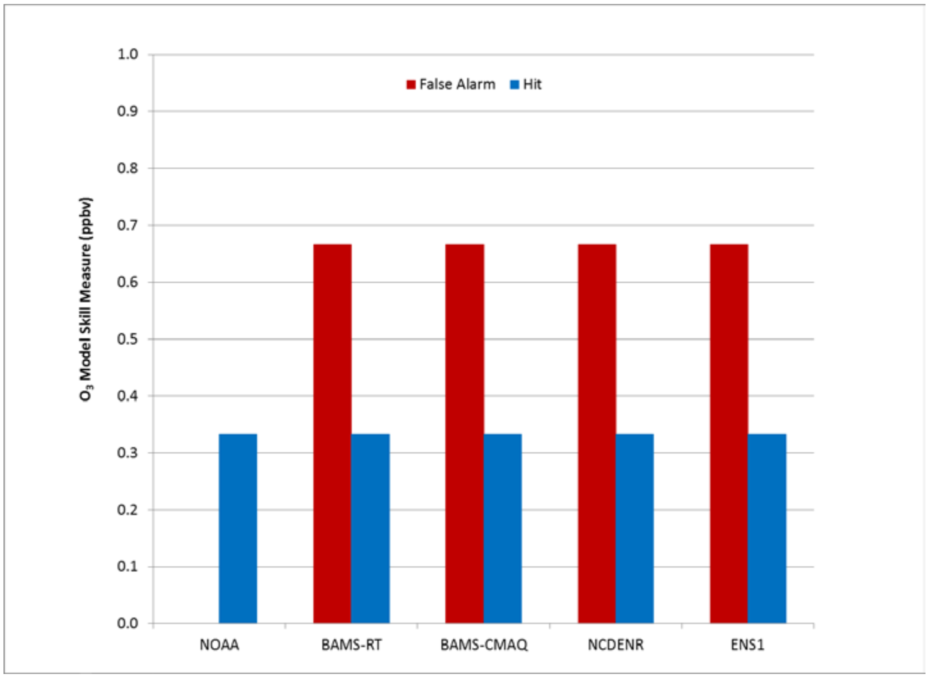

Figure 19. False alarm rate and hit rate for Code Orange O_3 predictions by air quality numerical forecast models in Delaware in 2014.

Appendix A. Skill Measures for Threshold Forecasts

The determination of the skill of a threshold forecast (e.g., Code Orange air quality) begins with the creation of a contingency table of the form:

Contingency Table for Threshold Forecasts			
		Observed	
		Yes	**No**
Forecast	**Yes**	a	b
	No	c	d

For example, if Code Orange O_3 concentrations are both observed and forecast ("hit"), then one unit is added to "a." If Code Orange O_3 is forecast but not observed ("false alarm"), then one unit is added to "b."

Basic Skill Measures

A basic set of skill measures are determined and then used as the basis for further analysis.

$$\textbf{Bias (B)} = \frac{a+b}{a+c}$$

Bias determines whether the same *fraction* of events are both forecast and observed. If B = 1, then the forecast is unbiased. If B < 1 there is a tendency to under-predict and if B > 1 there is a tendency to over-predict.

$$\textbf{False Alarm Rate (F)} = \frac{b}{a+b}$$

This is a measure of the rate at which false alarms (high O_3 forecast but not observed) occur.

$$\textbf{Hit Rate (H)} = \frac{a}{a+c}$$

The hit rate is often called the "probability of detection"

$$\textbf{Miss Rate} = 1 - H$$

Correct null forecasts:

$$\textbf{Correct Null (CNull)} = \frac{d}{c+d}$$

Accuracy:

$$\textbf{Accuracy (A)} = \frac{a+d}{a+b+c+d}$$

Other Skill Measures

Generalized skill scores (SS_{ref}) measure the improvement of forecasts over some given reference measure. Typically the reference is persistence (current conditions used as forecast for tomorrow) or climatology (historical average conditions).

$$\textbf{Skill Score } (\mathbf{SS_{ref}}) = \left(\frac{A - A_{ref}}{A_{perf} - A_{ref}}\right) * 100\% = \text{ nn\%}$$

The skill score is typically reported as a percent improvement of accuracy (A) with respect to a reference forecast. The reference forecast accuracy (A_{ref}) is typically climatology or persistence. The perfect forecast (A_{perf}) is usually 1 (e.g., for hits) or 0 (e.g., for false alarm).

Additional measures of skill can be determined. The Heidke skill score (HSS) compares the proportion of correct forecasts to a no skill random forecast. That is, each event is forecast randomly but is constrained in that the marginal totals (a + c) and (a + b) are equal to those in the original verification table.

$$\textbf{HSS} = \frac{2(ad - bc)}{(a+c)(c+d) + (a+b)(b+d)}$$

For this measure, the range is [-1,1] with a random forecast equal to zero.

Another alternative is the **critical success index (CSI) or the Gilbert Skill Score (GSS)** also called the **"threat" score**.

$$\textbf{CSI} = \frac{a}{a+b+c} = \frac{H}{1+B-H}$$

For this measure, the range is [0,1]. Since the correct null forecast is excluded, this type of measure is effective for situations like tornado forecasting where the occurrence is difficult to determine due to observing bias, i.e., tornados may occur but not be observed. This can also be the case for air quality forecasting when the monitor network is less dense. Note, however, that the random forecast will have a non-zero skill.

The **Peirce skill score (PSS), also known as the "true skill statistic"** is a measure of skill obtained by the difference between the hit rate and the false alarm rate:

$$\textbf{PSS} = \frac{ad - bc}{(a+c)(b+d)} = \text{H} - \text{F}$$

The range of this measure is [-1,1]. If the PSS is greater than zero, then the number of hits exceeds the false alarms and the forecast has some skill. Note, however, that if d is large, as it is in this case, the false alarm value (b) is relatively overwhelmed. The advantage of the PSS is that determining the standard error is relatively easy.

References

Stephenson, D. B., Use of the "odds ratio" for diagnosing forecast skill, *Wea. Forecasting*, **15**, 221-232, 2000.

Wilks, D. S., *Statistical Methods in the Atmospheric Sciences*, Academic Press, 467 pp., 1995.